NBER WORKING PAPER SERIES

# TOWARD AN UNDERSTANDING OF THE ECONOMICS OF APOLOGIES: EVIDENCE FROM A LARGE-SCALE NATURAL FIELD EXPERIMENT

Basil Halperin
Benjamin Ho
John A. List
Ian Muir

Working Paper 25676
http://www.nber.org/papers/w25676

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2019

This manuscript was not subject to prior review by any party, as per the research contract signed at the outset of this project. The views expressed here are solely those of the authors. List was Chief Economist at Uber when the research was completed, and Halperin and Muir were economists on the Ubernomics team. List and Muir are currently economists at Lyft. Thanks to Liran Einav and three anonymous reviewers for their insights, seminar participants at AFE 2017, AEA 2018, and Williams College, and AEA 2019; discussant Chiara Farronato; and to Courtney Rosen; for helpful comments and assistance. AEA Registry number: AEARCTR-0002342. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Toward an Understanding of the Economics of Apologies: Evidence from a Large-Scale Natural Field Experiment
Basil Halperin, Benjamin Ho, John A. List, and Ian Muir
NBER Working Paper No. 25676
March 2019
JEL No. C9,C93,D80,D91,Z13

**ABSTRACT**

We use a theory of apologies to design a nationwide field experiment involving 1.5 million Uber ridesharing consumers who experienced late rides. Several insights emerge from our field experiment. First, apologies are not a panacea: the efficacy of an apology and whether it may backfire depend on how the apology is made. Second, across treatments, money speaks louder than words – the best form of apology is to include a coupon for a future trip. Third, in some cases sending an apology is worse than sending nothing at all, particularly for repeated apologies. For firms, *caveat venditor* should be the rule when considering apologies.

Basil Halperin
Department of Economics
Massachusetts Institute of Technology
Cambridge, MA 02141
basilh@mit.edu

Benjamin Ho
Department of Economics
Vassar College
124 Raymond Ave
Poughkeepsie, NY 12604
benjaminho@gmail.com

John A. List
Department of Economics
University of Chicago
1126 East 59th
Chicago, IL 60637
and NBER
jlist@uchicago.edu

Ian Muir
Lyft
muir.ian.m@gmail.com

"Virtually every commercial transaction has within itself an element of trust... It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence." Arrow (1972)

# 1 Introduction

Economists have come to recognize the importance of trust, reciprocity, and other *social preferences* for explaining human behavior: people are self-interested, but also are often concerned about the payoffs of others (e.g., Rabin (1993), Charness and Rabin (2002), Fehr and List (2004)). Additionally, as Arrow (1972) and Sen (1977) have argued, networks of trust and reciprocity are essential for undergirding all economic exchange. However, relatively less is known about the consequences of *violations* of trust or reciprocity. What actions can be taken to avoid the deterioration of mutual confidence when trust has been compromised?

One common action to avoid the collapse of a relationship after a violation of trust or an unfortunate incident is to deliver an apology. The act of apology is an important thread running through households, friendships, and employer-employee relationships. Recent research has lent important insights into apologies in lab contexts and small-scale field experiments (see, e.g., Gilbert et al. (2017), Ho (2012)), but much remains ill-understood. For instance, why do firms apologize? Do customers actually value apologies?

With these questions as motivating examples, we begin by outlining a principal-agent model of trust violation and apologies. In the model, a customer (the agent) purchases output that provides a noisy signal of the underlying trustworthiness of a firm (the principal). Depending on the stochastic quality of the output, the firm may choose to apologize by sending a (potentially costly) signal to the consumer in an attempt to signal trustworthiness and restore the relationship. Several insights emerge from the model: among them, (1) in order for the apology to be an effective signal, it must be accompanied by a real cost; (2) the apology may backfire, i.e. in some circumstances, apologizing may be worse than not apologizing; (3) the efficacy of an apology depends on the familiarity of the consumer with the service; and (4) an apology has a greater effect on firm services that are dissimilar to services that the consumer normally consumes (Ho, 2012).

We leverage our theory to design a field experiment on the Uber ridesharing platform, which is a natural setting to lend insights into the underpinnings of the model. Uber is concerned that inaccurate estimates of trip duration may lead to decreased trust in the platform and decreased spending in the Uber marketplace. Because Uber services 15 million rides each day (Bhuiyan, 2018), even an extremely small fraction of rides being late could have large repercussions. Indeed, our analysis suggests that, absent any apology, a rider who experienced a late trip spends 5-10% less on the platform relative to a counterfactual rider, suggesting that there are material consequences to precisely the breach of trust described above.

With this substantial loss in revenues as a backdrop, we conduct the first large-scale, natural field experiment to measure the importance of apologies as a method for restoring trust in a relationship. In doing so, we design the experiment to have a tight link with the theoretical model. Our experiment is conducted across the United States over several months, sending real-time apology emails following a late trip, as defined by the actual trip time compared to the initial time estimate shown to the rider. We combine our experimental variation with rich customer data from Uber, the customer-firm relationship history, and situational context to test the specific predictions of the model.

A key goal is to measure the role of apologies in maintaining relationships with customers who have received a bad trip experience, measured by the level of future spending with the firm, and then to unpack the mechanisms through which apologies operate. The main set of treatments varies whether a customer receives an apology, the type of apology, and the size of the promotional coupon the customer receives as part of that apology ($5 or zero). We complement these treatments with a secondary set of treatments that send up to two additional apologies following a second and third delayed trip.

We report several interesting insights. First, a costly apology after a bad ride – in the form of a $5 coupon for a future trip – is an effective signal that increases future demand for future trips. Alternatively, we find that a signal in the form of an apology without a promotion (i.e. words alone) had little effect or was even sometimes counterproductive. As a placebo check, we find that the \$5 coupon administered directly after a bad ride is more effective than a \$5 coupon administered at a random time and unrelated to a rider's experience. We also find that the benefit of a costly apology can be detected even three months after the initial bad experience, whereas any benefit from a non-pecuniary apology quickly fades. This is especially notable because we measure the benefit as net of the coupon cost.

Second, we consider two other mechanisms suggested by theory that potentially create non-pecuniary costs of apologies, i.e. costs incurred by the firm besides a direct payment or coupon. We find that one additional cost is the potential for apologies to backfire, in particular when the apology included a promise to do better in the future. Our data suggest that in these cases repeated apologies after several bad experiences make things worse relative to fewer apologies. Apologies can restore trust but consumers who receive an apology hold firms to a higher standard in the future. If that future expectation is violated, apologies backfire.

The other possible non-pecuniary cost of an apology suggested by theory is the possibility that an apology could reduce demand for some kinds of rides while increasing demand for others. For example, consider a consumer who cares about two dimensions of quality: first, the ability to get quickly to the airport, and second the firm's overall customer service. An apology after a ride to the airport could serve as an admission of incompetence in providing airport rides, but cause the consumer to have more favorable beliefs about the firm's customer service. Thus, we would expect the apology to cause the consumer to increase the number of non-airport rides that they take. Unfortunately, the data lacked the power to make a conclusive statement about this channel.

Finally, we find that characteristics of trips and individuals affect the impact of apologies. The efficacy of an apology depends on the severity of the unsatisfactory service – in this case measured by how late the ride was, in minutes. In particular, we find a U-shaped relationship between severity of the unsatisfactory experience and apology effectiveness: for slightly bad quality and severely poor experiences, apologies are effective. Yet, for moderately poor experiences, apologies are not as effective. Moreover, the efficacy of an apology critically depends on a user's familiarity with the service. Apologies are less effective for users who are quite familiar with the product, yet are much more effective when the user has less experience with the Uber product. Both of these results are in concert with our model.

Our study fits in nicely with several strands of related work. First, it extends the social preference literature into an area that considers how trust can be restored after it is compromised. As Levitt and List (2007) summarize, lab and field experiments with the canonical trust game, dictator game, and other games have shown that the concepts of trust and reciprocity are essential for explaining human behavior. Rabin (1993), Charness and Rabin (2002), and Dufwenberg and Kirchsteiger (2004) formally model these concepts. Second, the extant literature on the economics of apologies has primarily been limited to small scale field and lab experiments (e.g. Aaker et al. (2004), Abeler et al. (2010), Fischbacher and Utikal (2010), Gilbert et al. (2017), Chaudhry and Loewenstein (2017)), or difference-in-difference analysis of policy interventions (e.g. Ho and Liu (2011)). We extend this literature by testing the model in the field, with detailed customer and situational data, and we follow the subjects for three months after the apology to measure how effects persist over time. Our data show that methodologically the lab studies have given us a key first look at the efficacy of apologies.

The remainder of our paper proceeds as follows. We first introduce the principal-agent model that guided the experimental design. Then we provide details of the experimental design, briefly describe the Uber ridesharing platform, and discuss the empirical results. We conclude with a discussion exploring how firms and individuals can use our results to further their understanding of apologies.

## 2 Theoretical Motivation

Our theoretical framework is based on the Ho (2012) principal-agent model of a customer-firm relationship that formalizes many of the findings about apologies in the psychology literature.[1] The model is a two-player game between a firm (the agent) and a consumer (the principal). Firms can be a good "high" type (e.g. high trustworthiness) or bad "low" type (e.g. low trustworthiness), $\theta \in \{\theta_H, \theta_L\}$. The firm produces output $y$ for the consumer, generating utility for the consumer. The quality of the output – how long the ride takes to arrive to the destination relative to expectations in our case – depends on firm type $\theta$ as well as external circumstance, $\omega \in \Omega$, that is uncorrelated with firm type (e.g. unexpected weather). Bad outcomes (i.e. low-quality output) can result from a firm with bad intentions, $\theta = \theta_L$, or alternatively from a bad draw from the state of nature $\omega$. The consumer is only aware of the overall quality of output $y = y(\theta, \omega)$. We can think of the firm's *intent* as the expected output over all possible external circumstances, $\omega$ which the firm does not know in advance, holding the firm's actual type, $\theta$, fixed: $E_{\hat{\omega}\in\Omega} y(\theta, \hat{\omega})$. The type, $\theta$, is known to the firm but unknown to the consumer. Type is defined so that higher types have "better" intentions. We call $\theta$ intentions because it represents expectations; even high-type firms may have a poor realization in any particular interaction.

There may be many dimensions of quality over which a firm may wish to signal their competence. For example, depending on the context and the particular consumer, higher quality could mean better on-time performance, or more responsive customer service, or something else entirely. What all these dimensions have in common is that higher quality represents higher expected future utility for that particular customer. We let $\theta$ represent any dimension of quality that yields higher expected payoffs for a consumer relative to their outside option. Formally, $\theta$ is defined as a match quality parameter that is pos-

[1]For example in lab experiments, Ohtsubo et al. (2012) and de Cremer et al. (2011) find that costly apologies can work better than cheap apologies; Skarlicki et al. (2004) and Kim et al. (2004) find that apologies can backfire; and many find that the efficacy of an apology depends on the type of offense (e.g. Maddux et al. (2011)).

itively correlated with consumer's expected utility due to a supermodularity condition. Ho (2012) has examples of how this supermodularity assumption can accommodate specific functional form assumptions about what quality could represent, such as lower cost of effort or greater concern for the principal's welfare (i.e. altruism).

Within the context of the rideshare industry, the timeline of the baseline game proceeds as follows (Figure 1). The consumer begins with a prior $p$ on the probability that the firm is high type. She then experiences a good or bad outcome for a ride, $y(\theta, \omega) \in \mathbb{R}$. Next, the firm chooses to apologize or not $a \in \{0, 1\}$. Finally, given the quality of the ride $y$ and apology or non-apology $a$, the consumer updates her beliefs about the firm's type, learns that an outside option is of high type with probability $p_{out}$, and then chooses to stay with the firm or to go with the outside option.

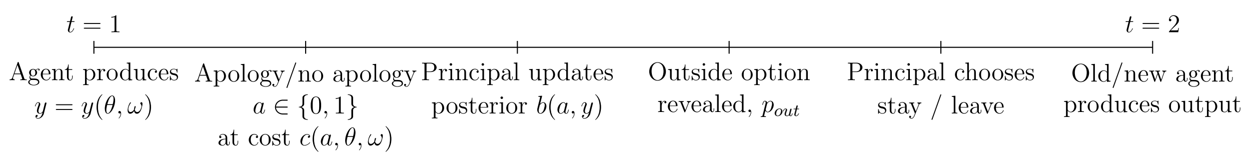



Figure 1: Timeline of the Apology Game

The consumer cares only about maximizing her consumption of rides, where ride quality $y(\theta, \omega)$ is a function of the firm's type $\theta$ and external circumstances $\omega$ such as traffic or weather. The consumer's choice, $x$, is simply whether to purchase from the same rideshare firm in period two or to take an outside option (e.g. switch to a competitor or take public transit).

$$U_{consumer}(x) = \sum_{t=0,1} y(\theta_t(x), \omega_t)$$

To keep things simple for this application, the firm's problem is simply to decide whether or not to apologize. As in Ho (2012), we abstract away from the firm's choice of effort in the determination of output quality. Under fairly general assumptions, specifically that choice of effort is supermodular with respect to firm's type $\theta$, a model that includes costly effort is equivalent to a reduced form model where cost of effort is subsumed into an apology cost function that depends only on type, $\theta$. The firm receives a profit per customer of $\pi$ and pays the apology cost (potentially zero) given by $c(a|\theta, \omega_1)$, which can depend on its type and the state of nature.

$$U_{firm}(a) = \pi \cdot x - c(a|\theta, \omega_1)$$

For the moment, assume the cost of apologies is constant: $c(1|\theta, \omega_1) = \kappa$. We will discuss other cost functions and cheap apologies (i.e. $c(1|\theta, \omega_1) = 0$) below.

Given this simple framework, the consumer observes signals about the firm's type, $\mathcal{H}$ , which in this case includes the firm output $y$ and the firm apology $a$. The consumer chooses to stay with the firm provided their posterior belief, given by $b(\mathcal{H}) \equiv Pr[\theta = \theta_H|\mathcal{H}]$, is greater than the quality of the outside option: $b(\mathcal{H}) > p_{out}$. The quality of the outside option is drawn from some known distribution $F(\cdot)$. The firm chooses to apologize if and only if:

$$\pi \cdot [F(b(y, 1)) - F(b(y, 0))] > \kappa$$

The efficacy of an apology, $\Delta b \equiv b(a = 1) - b(a = 0)$, is the impact the apology has on the customer's beliefs (i.e. the firm's reputation) and thus the likelihood that the customer will stay with the firm. The model provides several useful predictions about apology efficacy, $\Delta b$, that inform our experiment. Below, we discuss how apology efficacy is affected by uncertainty, the costliness of the apology, and the severity of the bad outcome. We also discuss predictions regarding repeat apologies.

## 2.1 Role of Uncertainty and the Role of Costs

A separating equilibrium where apologies signal higher type exists given the usual single crossing conditions: From Proposition 2 in Ho (2012), there are three existence conditions that allow a separating equilibrium to exist: 1) it is cheaper for high types to apologize, 2) continuing the relationship is more beneficial for high types, or 3) high types fail in different situations than low types. In the case of a repeated customer relationship, the second condition is most likely to hold as repeat customers will ultimately learn the firm's type just from repeat experience with the product. Therefore, the continuation value is lower for low quality firms since customers will eventually discover they are inferior and switch to the outside option. Accordingly, high types are more likely to maintain a lasting relationship. We examine the data for evidence for the other two existence conditions by exploring the value of implicit promises and the role of situation on the efficacy of the apology. We return to these questions in the Discussion.

In a separating equilibrium, three properties about the efficacy of an apology follow straightforwardly from Bayes Rule (see Ho (2012) for details):

1. Apologies are more effective when there is greater uncertainty in the relationship (when the prior $p$ is bounded away from 0 or 1)
2. Apologies are more effective early in relationships
3. Apologies are more effective the greater the apology cost. Further, apologies are only effective when there is a cost ($c(a = 1) > 0$)

Property 1 comes from the fact that when the prior belief, $p$, about the firm's type is close to 0 or close to 1, then the posterior belief is unlikely to change much given a single additional signal (the apology) and therefore the apology is likely to be ineffective. Apologies move beliefs the most when the customer is most uncertain. Property 2 follows from Property 1. A customer receives more and more signals about a firm's type over time. As the history of signals, $\mathcal{H}$, lengthens, beliefs converge to either 0 or 1. Therefore, apology efficacy is greater early in a relationship.

Finally, Property 3 is based on the cost of apologies. If apologies increase reputation then all firms will want to apologize. If costs are too low, then all types of firms will apologize. If all firms apologize with the same frequency then the efficacy of apologies tends toward zero. Apologies need to be costly in order to ensure good firms and bad firms apologize at different rates, which creates the separation in beliefs necessary for apologies to function.

## 2.2 Severity of Outcomes

We can apply the above results to also make predictions about how the efficacy $\Delta b$ of an apology varies according to the outcome $y$. Apologies are more effective when there is greater uncertainty about the firm's type. This is why we don't see apologies after good outcomes. Presumably people choose firms they have a good impression of. It is only when a bad outcome causes the customer to question that impression, that an apology would be justified to mend that impression. By a similar logic, we expect apologies to less effective for moderate lateness, than for extreme lateness.

Consider the distribution of possible outcomes (as measured by minutes late) for a firm with good intentions $\theta_H$ versus a firm with bad intentions $\theta_L$. Here we suppose that the lateness of a trip is given by a normal distribution, with a lower mean for high-type firms than for low-type firms, and common variance (Figure 2).

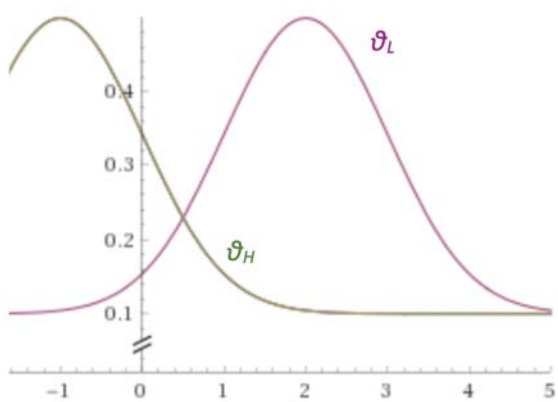



Figure 2: Distribution of Lateness of Outcomes (minutes)
Assume lateness of outcome is normally distributed, with high-type firms having a lower mean then low-type firms, and common variance (e.g. weather or traffic).

In this example, certainty that the firm's intentions are bad is maximized at the mean of the $\theta_L$ distribution. There is more uncertainty when the ride is less late since the firm is more likely to have had good intentions. Similarly when the ride is more late, the lateness is more likely to be due to the common shock (e.g. weather or traffic). As a result we would predict apologies to be least effective for intermediate values of lateness and more effective when barely late or extremely late.

## 2.3 Repeated Apologies

It is also useful to apply the above theory to make predictions regarding the efficacy of repeated apologies. Repeat apologies should be less and less effective as the customer gains experience with the firm. The customer is acquiring more and more information, and therefore is becoming more certain about the firm's type. Therefore the efficacy of an apology should diminish with increased interaction with the firm. In fact, Ho (2012) predicts that an apology could even begin to backfire if we assume apologies imply a promise for better behavior.

A cheap talk model of repeat apologies can lead to a backfire effect if we believe that an apology implies an implicit promise to do better in the future and repeated failure breaks that promise (as seen in the trust game experiment by Schweitzer et al. (2006)). A promise kept signals higher firm quality while a promise broken is worse than no apology at all. This can be seen in a simple screening contract extension to the baseline model.

Imagine the principal (consumer) offers the agent (firm) a menu that says the following: If the firm apologizes, then the relationship will be continued; however, if the firm is late again, the relationship will be immediately terminated in favor of the outside option. A separating equilibrium exists where good-intention firms apologize and accept the threat of immediate termination while bad-intention firms do not apologize and are judged in the future solely based on their performance (See Ho (2012) Online Appendix for details). In the context of Charness and Dufwenberg (2006) a broken promise signals lack of guilt aversion which serves as a second negative signal about the firm's type.

## 2.4 Heterogeneous Ride Types

Finally, consider the possibility that the firm offers different types of products (e.g. airport rides, weekend rides, rush hour rides). After a bad experience, consumers are uncertain whether the bad experience was due to the firm being bad overall or simply bad for that particular product. If an apology is seen as an admission that the firm is bad at that kind of ride, it could increase the consumer's impression of the firm overall. We would predict that a consumer who received an apology would be less likely to take rides similar to the bad one but more likely to take dissimilar rides from the same firm (See Ho and Huffman (2006) Online Appendix for details).

## 2.5 Hypotheses

In sum, the hypotheses from the model that are applicable to our setting include:

**Hypothesis 1** *The efficacy of an apology is higher when apologies are more costly.*

**Hypothesis 2** *The efficacy of an apology is lowest for intermediate severities of adverse outcomes when the variance of outcomes within types exceeds the variance of outcomes between types.*

**Hypothesis 3** *The efficacy of an apology is higher early in a customer-firm relationship, when there is greater uncertainty about the firm's type.*

**Hypothesis 4** *The efficacy of an apology decreases with repeated use and can backfire if overused.*

**Hypothesis 5** *An apology decreases future demand for similar trips but increases future demand for dissimilar trips.*

The model defines apology efficacy as the change in beliefs, $\Delta b$ that arise in response to an apology. While we do not observe the beliefs of our experimental subjects, we do observe their future decision of whether to stay with the firm, or to choose an outside option: $Pr\left[b(\mathcal{H}) > p_{out}\right]$. It is this outcome variable that we will use to test our main hypotheses.

# 3 Experimental Design

To test the hypotheses from this model, we conducted a natural field experiment (see Harrison and List (2004)) on the Uber ridesharing platform. The Uber platform connects riders with drivers willing to provide trips at posted rates. A rider provides her desired pickup and dropoff location through a phone app, and is offered a price, an estimated time to pickup, and an estimated time to destination (ETD). She then may choose to request an Uber ride and will be picked up and transported to the destination. At the end of the trip, the rider has the option to tip the driver (see also Chandar et al. (2018), Chandar et al. (2019)). This describes the standard "UberX" product offering which is the focus of our experiment, but Uber offers products that slightly vary this experience. For example, UberPOOL offers a discounted price but may involve trip detours to pick up multiple riders traveling along a similar route.[2]

One measure of platform quality is the accuracy of the ETD provided to riders. Rideshare firms such as Uber are justifiably concerned that inaccurate such estimates may lead to decreased trust and consequently decreased spending with Uber. As mentioned above, we completed an analysis using a matching methodology to identify the causal effect on future spending of a rider who experienced a late trip – a "bad ride" – relative to a statistically identical customer who took an identical ride but which arrived on time. This analysis, which helped to motivate the present study, found that riders in the right tail of the lateness distribution spend 5-10% less on the platform relative to the counterfactual. These results are available upon request.

To attenuate the costs of bad trips and to test the power of apologies, we designed a natural field experiment. Our field experiment was conducted over the course of several months in 2017. We selected six of Uber's largest markets to ensure a mix of cities with differing levels of competition between Uber and competing ridesharing platforms, and separately to ensure large enough

[2]Cohen et al. (2016) also use Uber data to study the demand side of the ridesharing market. A number of other papers use Uber data to examine the supply side, see e.g. Cook et al. (2018), Hall et al. (2017).

ridesharing markets to generate a sufficient sample size. 1.5 million subjects passed through the experiment across the eight treatment groups described below.

Riders entered the treatment upon experiencing a bad ride, defined as an UberX trip which arrived at the destination $n$ minutes later than the ETD initially displayed to riders when choosing whether to request a trip. The threshold $n$ varied by city based on the city's historical distribution of lateness. The threshold was set so that in expectation only the 5% latest trips would be classified as late in each city, which generally implied a 10-15 minute threshold.

An hour after the end of a bad ride, a customer in a treatment group would receive an email, the content of which varied depending on the treatment group. We then follow all of the customer's future interactions with Uber for 84 days.

Following our theory, subjects were divided among eight treatment groups (Figure 3). Half received a $5 promo code while the other half received no promo code. The promo code conditions were crossed with four different apology types:

1. No apology.
2. Basic apology: e.g., "Oh no! Your trip took longer than we estimated."
3. Status apology: e.g., "We know our estimate was off."
4. Commitment apology: e.g., "We're working hard to give you arrival times that you can count on."

The wording of each email was in the spirit of our model and followed Ho (2012). The different kinds of apologies were designed to emphasize different apology mechanisms. In particular, the "Status" apology was designed to amplify the effect of apologies on dissimilar rides, and the "Commitment" apology to emphasize the effect on repeated failures. The messages were sent as emails, with subject lines that suggested the nature of the apology and highlighted the $5 promotion if attached. Full message details along with the theoretical motivations for each apology type are found in Appendix A and B.

Treatment groups were balanced on eight dimensions:

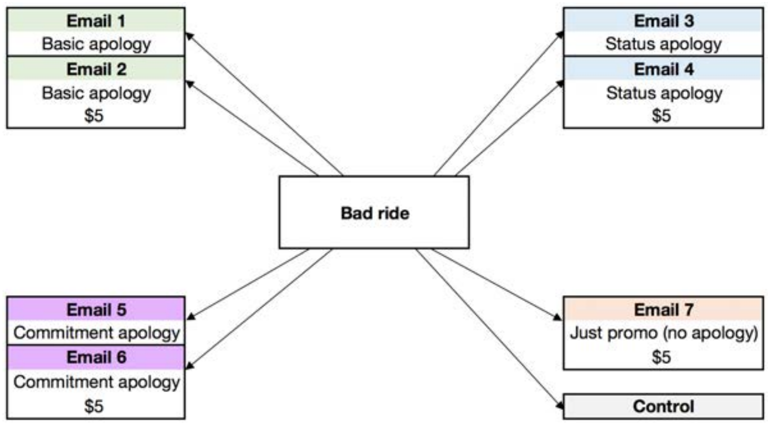



Figure 3: Treatments

The experiment was a 4x2 design with 4 apology message types crossed with either a no promo code condition or a $5 coupon condition.

1. Average fare previously faced by a rider (in all of 2017 prior to the experiment launch)
2. Days since signing up with Uber
3. Lifetime dollars spent on Uber (up until experiment launch)
4. Lifetime trip count (up until experiment launch)
5. (Number of UberPOOL trips taken in life)/(Number of UberX + UberPOOL trips taken in life)
6. Number of UberPOOL trips taken (in the month before experiment launch)
7. Number of UberX trips taken (in the month before experiment launch)
8. Number of support tickets filed (in all of 2017 prior to the experiment launch)

Technological limitations meant balancing could only be done for subjects who had signed up for the Uber platform before the start of the experiment. Subjects who joined after the start date were randomly assigned to one of the treatment groups. As a result, because of the large number of subjects, means were significantly different in t-tests between groups, but the differences were economically small, as reported in Table 1. Appendix B contains further details on experimental design, including the language and imagery contained in the apology email.

Table 1: Balance Check – Mean Rider Characteristics by Treatment

| | avg_fare | days_since_signup | lifetime_billings | lifetime_pool_share | lifetime_trips | n_recent_pool_trips | n_recent_x_trips | n_tix |
|---|---|---|---|---|---|---|---|---|
| Control | -14.339 | 762.622 | 1990.844 | 0.124 | 131.44 | 1.277 | 5.363 | 0.896 |
| Basic apology | -14.318 | 757.443 | 1973.183 | 0.124 | 130.023 | 1.249 | 5.254* | 0.877 |
| Basic apology + promo | -14.366 | 761.054 | 1984.287 | 0.123 | 131.039 | 1.27 | 5.317 | 0.883 |
| Commitment apology | -14.276 | 759.031 | 1963.447 | 0.124 | 129.578 | 1.317 | 5.289 | 0.87 |
| Commitment apology + promo | -14.383 | 757.146 | 1979.742 | 0.123 | 129.618 | 1.242 | 5.279 | 0.863 |
| Status apology | -14.309 | 757.866 | 1974.789 | 0.122 | 129.4 | 1.234 | 5.222*** | 0.87 |
| Status apology + promo | -14.356 | 761.665 | 1995.218 | 0.124 | 131.619 | 1.28 | 5.377 | 0.893 |
| Just promo | -14.368 | 762.392 | 1994.212 | 0.124 | 131.528 | 1.281 | 5.351 | 0.886 |

* indicates significance of pairwise t-test versus the control group at the 5% level, with the Bonferroni correction applied. ** indicates the same at the 1% level and *** at the 0.1% level.

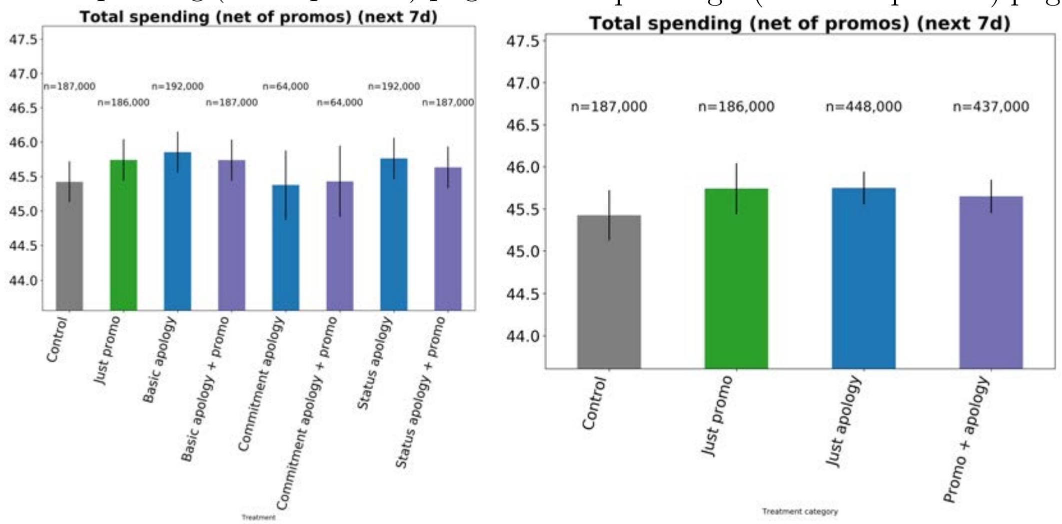
In general we report results for future spending net of any promotions applied ("net spending"), including but not limited to our \$5 promo. For example, if a rider took a single \$8 trip in the seven days following treatment, but used a \$5 promotion on that trip, her level of spending would be reported as \$3. The analysis using gross spending yields similar results. We also consider future trip count, future tipping, and the extensive margin of whether the rider took any future trips as outcome variables.

# 4 Results

We begin by presenting the unadjusted means of our main outcome variable, net spending, across the seven treatment groups versus the control group. Figure 4a presents average spending by riders over the seven days following the bad ride. The figure can be read as follows: we have 186,584 customers in the control group who had a bad trip. On average, these customers spent (net of promotions) \$45.42 in the seven days after the bad trip. Comparing this to the basic apology group, which had 191,825 subjects, we find that those who received our basic apology spent \$45.86 in the seven days subsequent to a bad trip. This result is significant at the $p < 0.05$ level using a standard t-test of means.

Another finding in the raw data is that we find no statistically significant differences between the different message types in the raw means. F-tests show that mean spending within the set of three "Just apology" treatments were statistically indistinguishable (ANOVA $p = 0.27$), as was mean spending within the set of three "Promo + apology" treatments (ANOVA $p = 0.63$).

Total spending (net of promos).png Total spending (net of promos).png




(a) (b)

Figure 4: Mean spending by treatment group

Panel (a) presents raw mean spending (net of any promotions) by treatment arm. Panel (b) aggregates the results across message types into four treatment categories (since the content of the messages themselves was found to be insignificant), with shaded 95% confidence intervals.

Accordingly, for ease of comparison, we aggregate the treatments into four categories, shown in Figure 4b. The categories are: the control group, the treatment group that received just the \$5 promo code ("Just promo"), the three treatment groups that received just an apology email ("Just apology"), and the three treatment groups that received both a \$5 promo and an apology ("Promo + apology"). The figure shows similar insights as observed in the disaggregated data: coupons are an important promotional tool, and apologies alone work marginally.

To complement this visualization of the raw data, we provide Table 2, which reports summary statistics for the full set of outcome variables, again at the seven-day horizon. Note that the effect of treatments on trip count and whether a rider takes a future trip are consistent with the effect on spending. Additionally, the effects of different treatments on future tipping behavior are indistinguishable from zero.

Table 2: Means (Std Errs) by Treatment Category (7d)

| | Total spending (net of promos) | Trip count | Total spending (incl. promos) | Total tips | Took another trip |
|---|---|---|---|---|---|
| Control | 45.424 | 2.848 | 46.6 | 0.977 | 0.674 |
| | (0.152) | (0.009) | (0.154) | (0.008) | (0.001) |
| Just promo | 45.741 | 2.877 | 47.219 | 0.994 | 0.680 |
| | (0.154) | (0.009) | (0.156) | (0.009) | (0.001) |
| Just apology | 45.748 | 2.851 | 46.924 | 0.991 | 0.672 |
| | (0.100) | (0.006) | (0.101) | (0.006) | (0.001) |
| Promo + apology | 45.649 | 2.879 | 47.166 | 0.994 | 0.679 |
| | (0.101) | (0.006) | (0.102) | (0.006) | (0.001) |

*Note:* Outcome variables at a seven-day horizon are presented here, but data were collected at horizons up to and beyond 84 days after the initial bad ride.

To supplement the raw data observations, we conduct a series of regressions. Our main empirical specification regresses the outcome variables of interest for each subject $i$ on the set of eight treatment dummies indexed by $j$, controlling for the variables $\vec{X}$ on which we balanced in addition to city, date, and hour-of-week fixed effects:

$$ln(Outcome_i) = \sum_j \alpha_j \cdot Treatment_j + \vec{\beta} \cdot \vec{X}_i + \gamma_{city} + \delta_{date} + \eta_{hour} + \varepsilon_i \quad (1)$$

Regression results for the effect of apologies on net spending, estimated using this specification, are presented in Table 3. Each column estimates the treatment effect on net spending over progressively longer horizons (7, 14, 28, 56, 84 days). A main feature to note is that the apology by itself (without a promotion) has no statistically significant effect at conventional levels. In fact, while the effect of an apology is largely not significant, if anything the presence of the apology in and of itself has a negative effect over longer time horizons (56 to 84 days). Table 4 presents the same specification but with number of future rides as an outcome variable. It shows the same basic pattern, therefore we will focus our attention on net spending as the outcome variable.

Figure 5 plots the estimated coefficients on the treatment dummies from our main empirical specification estimated over the same horizons described above. We find persistent effects of treatments that include a promotion as far out as three months after the apology was sent.

One possible explanation for the persistence of the effect is intertemporal

Table 3: Log of future net spend by treatment group over the $N$ days after the bad ride

| | 7d | 14d | 28d | 56d | 84d |
|---|---|---|---|---|---|
| Basic apology | 0.007 | 0.001 | -0.005 | -0.008 | -0.009 |
| | (0.006) | (0.006) | (0.006) | (0.005) | (0.005) |
| Basic apology + promo | 0.015** | 0.012* | 0.015** | 0.011* | 0.009 |
| | (0.006) | (0.006) | (0.006) | (0.005) | (0.005) |
| Commitment apology | -0.002 | -0.004 | -0.013 | -0.016* | -0.014 |
| | (0.008) | (0.008) | (0.008) | (0.008) | (0.007) |
| Commitment apology + promo | 0.008 | -0.001 | 0.001 | -0.005 | -0.006 |
| | (0.008) | (0.008) | (0.008) | (0.008) | (0.007) |
| Status apology | 0.006 | 0.002 | -0.003 | -0.006 | -0.006 |
| | (0.006) | (0.006) | (0.006) | (0.005) | (0.005) |
| Status apology + promo | 0.013* | 0.008 | 0.011 | 0.011* | 0.011* |
| | (0.006) | (0.006) | (0.006) | (0.005) | (0.005) |
| Just promo | 0.015** | 0.007 | 0.005 | 0.005 | 0.008 |
| | (0.006) | (0.006) | (0.006) | (0.005) | (0.005) |
| Controls | X | X | X | X | X |
| City | X | X | X | X | X |
| Date | X | X | X | X | X |
| Hour | X | X | X | X | X |
| No. observations | 1257738 | 1257737 | 1257735 | 1257735 | 1257740 |

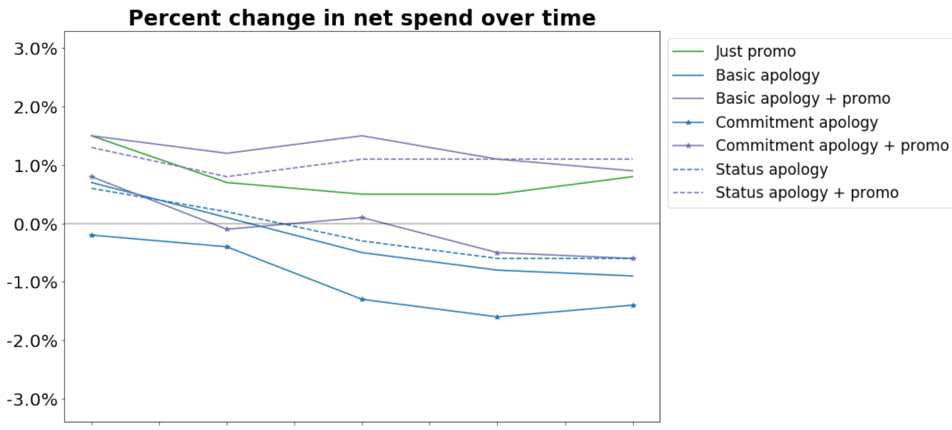
OLS regressions of log future net spending in the $N$ days after experiencing a bad ride with city, date, and hour-of-week fixed effects. Controls include: average fare; days since signup; lifetime billings; lifetime POOL share; lifetime trips; number of recent POOL trips; number of recent UberX trips; and number of support tickets filed.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

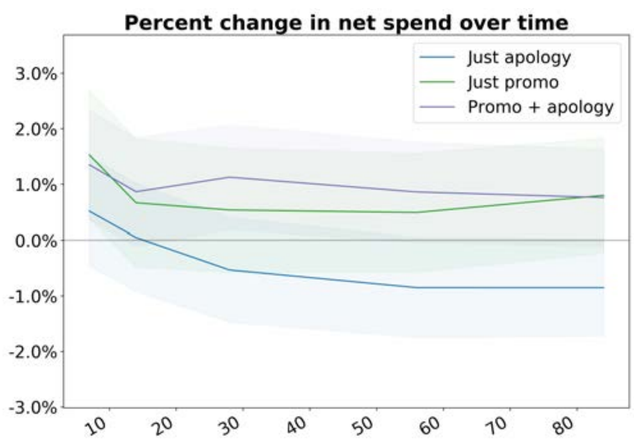
Table 4: Log of future number of rides by treatment group over the $N$ days after the bad ride

| | 7d | 14d | 28d | 56d | 84d |
|---|---|---|---|---|---|
| Basic apology | 0.004 | 0.001 | -0.003 | -0.005 | -0.006 |
| | (0.002) | (0.003) | (0.003) | (0.003) | (0.003) |
| Basic apology + promo | 0.011*** | 0.01*** | 0.01*** | 0.008* | 0.007* |
| | (0.002) | (0.003) | (0.003) | (0.003) | (0.003) |
| Commitment apology | -0.001 | -0.002 | -0.008* | -0.011* | -0.011* |
| | (0.003) | (0.004) | (0.004) | (0.005) | (0.005) |
| Commitment apology + promo | 0.008* | 0.003 | 0.002 | -0.001 | -0.002 |
| | (0.003) | (0.004) | (0.004) | (0.005) | (0.005) |
| Status apology | 0.003 | 2.68e-04 | -0.003 | -0.006 | -0.007* |
| | (0.002) | (0.003) | (0.003) | (0.003) | (0.003) |
| Status apology + promo | 0.01*** | 0.008** | 0.01*** | 0.01** | 0.01** |
| | (0.002) | (0.003) | (0.003) | (0.003) | (0.003) |
| Just promo | 0.01*** | 0.007* | 0.007* | 0.007* | 0.008* |
| | (0.002) | (0.003) | (0.003) | (0.003) | (0.003) |
| Controls | X | X | X | X | X |
| City | X | X | X | X | X |
| Date | X | X | X | X | X |
| Hour | X | X | X | X | X |
| No. observations | 1257788 | 1257788 | 1257788 | 1257788 | 1257788 |

OLS regressions of log future trips taken in the $N$ days after experiencing a bad ride with city, date, and hour-of-week fixed effects. Controls include: average fare; days since signup; lifetime billings; lifetime POOL share; lifetime trips; number of recent POOL trips; number of recent UberX trips; and number of support tickets filed.
*** p<0.01, ** p<0.05, * p<0.1




(a)




(b)

Figure 5: Percent change in spending over time

We plot the $\alpha$ coefficient on each treatment dummy from model (1), with total spending as the outcome, between the date of the bad ride and some future date 7, 14, 28, 56, and 84 days in the future. Panel (a) presents results for each treatment, and panel (b) aggregates the results across message types into four treatment categories for increased power.

complementarities in consumption. In other words, if taking an additional ride today increases a rider's chance of taking a ride tomorrow, then simply inducing a customer to take an additional ride in the first week could have persistent effects. While this result is intuitively appealing, it should be tempered in that if complementarities were the only force driving the persistence, one would expect the effect size to get smaller over time. In fact, if anything the effect (of a promotional coupon alone) stays steady or increases (albeit not significantly) by day 84.

What is especially notable in the results is that the effect of an apology by itself becomes more negative over time. An apology alone (with no coupon) becomes significantly negative by day 84, in contrast to the effect of an apology with a promo. In particular, the difference in effects of an apology without a coupon by day seven is statistically distinguishable from the effect by day 84 ($p < 0.001$), whereas the difference for the effect of an apology including a coupon is not ($p = 0.18$). These points confirm **Hypothesis 1** that apologies are more effective when the cost associated with the apology is higher.

Breaking the results down by treatment, we can see in Figure 5a that the downward time trend is seen primarily in the treatments with no coupon, along with both coupon and no-coupon treatments when a commitment was made. (The one treatment in this set that did not see a statistically significant decline was the commitment apology without a promo. However, the estimated coefficients for this treatment had a decline similar to the others, around 1.2%, with a p-value of 0.098. The lower significance for commitment apologies could be explained by the lower power in that treatment as we sub-divided the commitment treatment into 8 groups to measure repeat apologies, so the sample we are testing is only 1/8th as large as the other treatments.) There were declines in some of the other treatments but they were smaller and not statistically significant.

## 4.1 Heterogeneity by Severity of Lateness

Recall **Hypothesis 2** that an apology would be least effective for moderate levels of lateness, since this is when the poor experience is most likely attributable to the firm itself. On the other hand, apologies would be more effective for low levels of lateness (when the firm is more likely to be of the high type) and high levels of lateness (where the most severe delays can be attributed to external factors like weather).

This prediction is consistent with the pattern observed in the data. Figure 6 provides the estimated coefficient for the aggregated treatment variable interacted with indicators for the degree of lateness as measured by decile. Since there is significant variation in the distribution of lateness for each city, we measure lateness relative to other rides from the same city, although other specifications produce the same pattern. As predicted, apologies are least effective (or most damaging) for intermediate degrees of lateness.

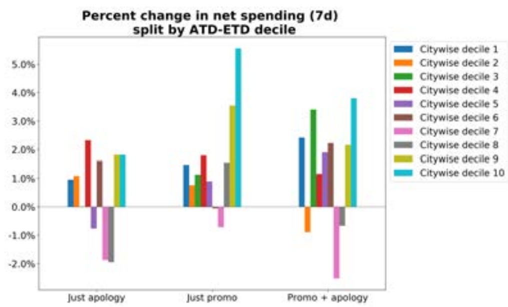



Figure 6: Efficacy of Apology by Severity of Outcomes

The coefficient on the treatment variable interacted with the decile of how late the ride was as measured by number of minutes relative to the other rides in the sample from the same city.

We test this relationship formally by estimating our main specification (1) with the addition of interaction terms between the treatment dummies and the percentile of lateness and the percentile squared. We find that the quadratic interaction term is statistically significant for the "promo + apology" treatment at the $p < 0.05$ level and for the "Just promo" treatment at the $p < 0.10$ level.

## 4.2 Heterogeneity by Rider History

We now turn to **Hypothesis 3** that apologies should be most effective when there is the greatest degree of uncertainty and therefore we would expect greater efficacy for new users of the ridesharing platform. Here we present the effect of apologies within subsamples of riders based on quartiles of riders' number of rides before having the bad experience.

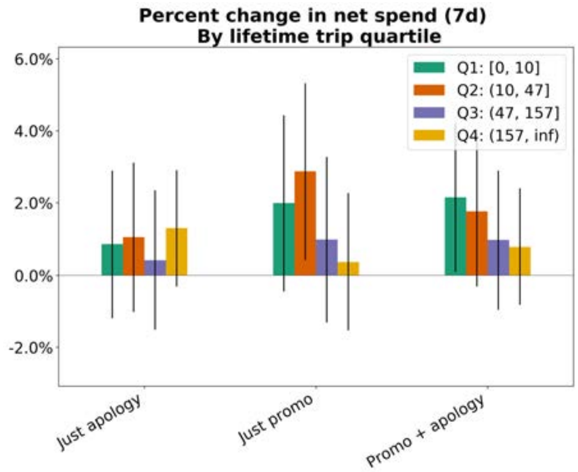



Figure 7: Rider heterogeneity

Treatment effect on net spending within subsamples defined by the quartile of the number of past rides in the customer's history.

As shown in Figure 7, our results are mixed. For the joint promo and apology treatment, the point estimates indicate that the treatment effect is highest for the newest quartile of users (those with 0 to 10 lifetime trips) and lowest for the most experienced users (those with greater than 157 trips), with the effectiveness decreasing across quartiles. However, for those who received just an apology, the point estimates are mixed, and in fact the treatment effect estimate is smallest for the newest users and somewhat higher for the most experienced users. In both cases, the confidence intervals are wide.

Looking instead at a different measure of unfamiliarity and uncertainty, the frequency of UberPOOL usage relative to UberX, we find results more consistent with the hypothesis (Figure 8). The two most popular services provided by Uber are UberPOOL and UberX. Since our experiment was conducted exclusively on UberX riders, we expect riders who have mostly used UberPOOL in the past to be more uncertain about the quality of UberX. Indeed, our point estimates indicate that riders who mostly used UberPOOL in the past were much more likely to be positively influenced by an apology than riders who mostly used UberX, although the confidence intervals are again large.

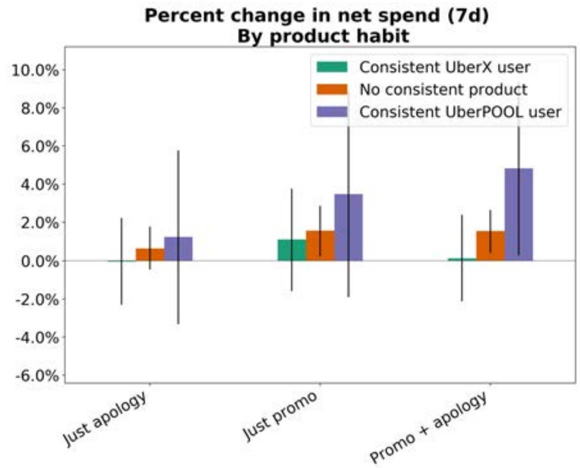


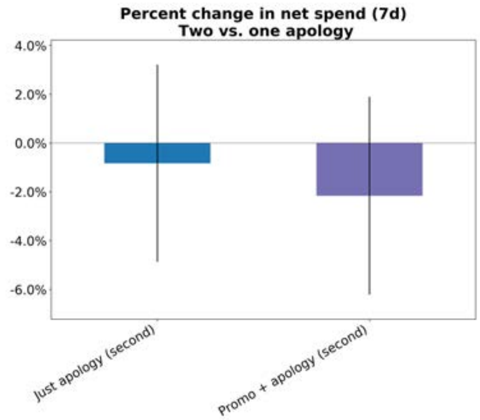

Figure 8: UberPOOL Riders vs UberX Riders

Treatment effect on net spending within subsamples defined by whether a rider is a "consistent UberX user" ($\geq$ 75% of trips in the preceding three months on UberX), a "consistent UberPOOL user" ($\geq$ 75% of trips in the preceding three months on UberPOOL), or the intermediate case with no consistent product.

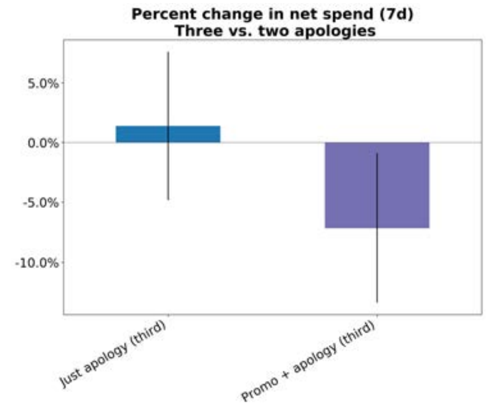
## 4.3 Impact of Repeat Apologies

Next, consider **Hypothesis 4**: that repeat apologies would be less and less effective over time and may even be counterproductive. For a subsample of riders we conduct the following secondary experiment. We split the sample and offer a second apology for half of the subjects who in following weeks receive a second late trip, leaving the other half as a control (having only received one apology). For the subsample who received two apologies we split the sample again for those who took a third late trip, offering half a third apology and leaving the other half as one final control (who only received two apologies).

As before, a cheap-talk apology alone without the $5 promotion remains largely ineffective. However, whereas the short term effect of the first apology with a $5 promotion yielded a 2% increase in net spending, the net effect on spending of the second apology is not significantly different compared to someone who had a second bad ride but received no new apology message. For the third bad ride, the apology on its own is insignificant again, while the third apology with a promotion has a significantly negative effect on future net




(a)




(b)

Figure 9: Treatment Effect of Repeat Apologies

Panel (a) reports the marginal treatment effect *over the first apology* of a second apology treatment for a second bad experience with Uber, compared to the relevant control group. Panel (b) reports the marginal effect of a third apology relative to the second. Effects are at a seven-day horizon.

spending relative to someone who received three bad rides but only received two apologies with a promotion. In fact, this negative effect shows up not just in terms of future net spending but also in terms of the number of future rides taken and in terms of future gross spending.

This backfire effect we observe is consistent with an apology acting as a promise. An apology can temporarily restore a customer's loyalty after an adverse outcome. However, an apology acts as a promise that the adverse outcome was due to unexpected external factors, and that the customer should therefore expect better outcomes in the future. When those higher expectations go unmet, the firm reputation suffers more than if no apology had been tendered at all. Apologies should therefore be used sparingly and ideally only after unexpectedly bad outcomes that are unlikely to repeat again in the near future.

### 4.4 Heterogeneity by Ride Type

Finally, consider **Hypothesis 5**: that an apology after a trip of a given type decreases demand for that category of ride but increases demand for dissimilar trips. Given the rich data associated with each trip on the Uber platform, we are able to classify trips into several natural categories based on popular Uber use cases. We consider the following categories of trips: rides to and from an airport; rides during morning commute hours; and rides during weekend hours. We also link trip timing and location to local weather conditions using the Dark Sky weather API and consider trips during times of precipitation (i.e. rain, snow, or sleet) versus those not during times of precipitation.

To test the model's hypothesis about heterogeneous effects due to the circumstance of the bad ride, we compared the treatment effect of apologies on riders who had (for example) a bad airport trip on future airport trips versus the treatment effect on future non-airport trips, However, we are unable to reject the null that apologizing has no differential effect between trip types for the categories tested (airport vs. non airport, rush hour vs. non rush hour, weekend v.s weekday, and rainy vs. non-rainy).

## 5 Discussion

Since our principal finding is that it is primarily a promotional coupon that can be used for a future trip that restores the firm's reputation and not the apology itself, one can ask: is this an "apology effect" or just a "promo effect"? One approach to answer this query is to compare our estimated effect sizes with the effect of a generic $5 promotion sent out randomly by Uber, which will have no apology connotation.

Running concurrently in the cities where our experiment was conducted (between the months of June and October of 2017), another experiment tested the effects of randomly sending a $5 promo against a control group that received no promotion. While this serves as an important comparison experiment, we should note that this natural field experiment is not a perfect analogue to our main apology experiment for two reasons. First, this experiment proactively targeted the entire Uber rider population whereas our own experiment targeted only those who had received a late ride. Having a late ride is more likely to happen to more frequent riders simply by chance: more trips implies a higher chance of at least one bad draw. To make treatment effects comparable, we restrict consideration to just those riders who experienced at least one bad ride during 2017. It is important to note that while these riders experienced a bad ride, the random $5 promos were not sent because of this ride and could have been sent months before (or after) the experience.

A second limitation of our comparison experiment is that this generic promo was usable multiple times and limited to a single week, whereas our promotion was one-time use in the next three months. Therefore, we might expect this generic promotion to be much more effective at the seven-day horizon than our apology promotion.

In fact, while the sample size is small ($n = 27,203$), we find that our "just promotion" treatment in the aftermath of a bad experience is statistically significantly more effective than the randomly-timed generic promotion. Stacking the generic promotion data with our "just promotion" and control data, we estimate:

$$ln(Outcome_i) = \alpha_1 \cdot is\_generic + \alpha_2 \cdot is\_treated + \alpha_3 \cdot is\_generic \cdot is\_treated + \vec{\beta} \cdot \vec{X}_i + \gamma_{city} + \delta_{date} + \varepsilon_i, \tag{2}$$

where the coefficient on the interaction term $\alpha_3$ is the treatment effect of receiving a generic (randomly-timed) promotion, compared to receiving a promotion in the aftermath of a late trip.

Estimating (2) with net spend as the outcome variable at the seven-day horizon, and using the same set of controls as in the previous analyses, we find that the randomly-timed promotion has a significantly negative effect of **-8.3**% (p-value $< 0.001$) on future net spending, which is in contrast to the positive effects of the $5 "just promotion" without an apology. Importantly, this suggests that it matters that the act of remediation occurred after an adverse event, a breach of trust. This is, at least, consonant with the idea that our \$5 "just promotion" treatment had an extra impact after a bad trip compared to the effect observed after a generic \$5 promotion is received.

This finding also lines up with the findings of an experiment that was run independently, and concurrently, by the ridesharing company Via (Cohen et al., 2018). This study also found that while a \$5 promo after a bad ride was effective at increasing net spending, a \$5 promo randomly given had an insignificant effect on gross and net spending. Cohen et al. (2018) also find that a cheap apology (without a promotional coupon) had no significant effect. This replication with a different company is encouraging in that it suggests our results generalize to rideshare firms beyond Uber, which had perhaps a unique reputation at the time our experiment was conducted. There are a couple differences observed between the Cohen et al. (2018) paper and our own that are worth noting. They find that apologies mostly matter for late pickups, whereas our experiment focused on late arrivals. Indeed they find null results for late arrivals. They also find that their apologies are most effective for their most frequent customers whereas we found indistinguishable treatment effects on users by frequency. These differences are likely due to Via's model which emphasizes shared rides. When a user hails a ride with Via, she knows that the driver will pick up other riders along the way. Thus, she does not necessarily have the same expectation for an on-time arrival.

The Via study, occurring in a different geography and different setting, is also informative because apologies are undoubtedly context-dependent. Abeler et al. (2010), who study apologies on an auction website, is similarly complementary. Interestingly, they find that cheap apologies were more effective than monetary compensation. We have two possible explanations for the incogruence between our results and Abeler et al.'s insights. First, their outcome variable was the customer's rating of the seller on the auction website. This is relatively costless for the customer to change. The second is that their offer of monetary compensation was offered as a quid pro quo payment to the customer to change the rating (which may have been construed as a bribe) whereas in our case, the monetary compensation was offered as a gift. Of course, our thoughts are merely speculative, and further experiments are needed to precisely identify the role of norms and context.

Future research can also better identify the mechanisms that determine how apologies work. Apologies can contain monetary restitution, admission of guilt, promises about the future, expression of empathy, or even excuses (See Appendix for more details). The experiment was designed to test different apology mechanisms by varying the message that accompanied the apology and by estimating the effect of apologies in different traffic and weather situations. While some of the effects of different apology messages were directionally consistent with predictions from theory, the significance of their effect was not consistently robust, perhaps because the email text associated with promotions was not carefully read by customers (email open rates averaged approximately 30%). Similarly, the efficacy of apologies did vary by weather and traffic, but not in any systematic way discernible through the lens of theory.

# 6 Conclusion

We present results from a large-scale natural field experiment on the effects of apologies to restore trust within a principal-agent relationship. We offer not just evidence that apologies matter for customers, but also insights into how apologies matter. Our results have implications both for firms deciding how, and when, to apologize and for understanding how trust can be repaired in economic relationships more generally.

We find that the most effective apology was the provision of a $5 coupon, with or without any accompanying apology text. Giving such a coupon after a bad ride was more cost-effective than $5 coupons given at random. Yet, we find that the benefits of apologizing with a coupon disappear after 3 months when a promise to do better in the future is made. We further examine dimensions of customer characteristics and characteristics of the adverse outcome that could help provide guidance for more effective apologies going forward, such as the customer's familiarity with the product. Furthermore, apologizing repeatedly to the same person who had multiple bad experiences in a three-month period actually reduced future spending, relative to someone who also had repeated bad rides but did not receive repeated apologies.

Overall, our experiment provides real world empirical support for the general apology model. While previous lab studies have served to provide important insights, our data demonstrate the value of the signaling view of apologies by showing that its predictions hold in the field. Our analysis also provides useful advice for firms on the ifs, whens, wheres, and hows to apologize optimally. We find that while apologies can be an effective way to restore and prolong the customer relationship, the reason why apologies are not more frequent is because they are costly and potentially backfire. Firms often do not apologize because apologizing is difficult. Our data highlight that the safest way to remediate a bad experience is a simple promotion applied to future purchases. We find that money spent in this way, after an adverse event, yields a positive return for the firm even when promotions sent at other times do not.

There are several opportunities to expand on our experiment. Future work should explore the impact of apologies in other industries and include greater variation in the cost dimension. In particular, we remain interested in exploring the role of different kinds of apologies where the implicit promises associated with an apology are made even more explicit.

# References


Aaker, J., Fournier, S., and Brasel, S. A. (2004). When Good Brands Do Bad. *Journal of Consumer Research*, 31(1):1–16.

Abeler, J., Calaki, J., Andree, K., and Basek, C. (2010). The power of apology. *Economics Letters*, 107(2):233–235.

Arrow, K. J. (1972). Gifts and Exchanges. *Philosophy & Public Affairs*, 1(4):343–362.

Battaglini, B. Y. M. (2002). Multiple Referrals and Multidimensional Cheap Talk. 70(4):1379–1401.

Bhuiyan, J. (2018). Uber powered four billion rides in 2017. It wants to do more – and cheaper – in 2018.

Chandar, B., Muir, I., List, J. A., and Gneezy, U. (2018). The Determinants of Tipping: Evidence from Rideshare. *Working paper*.

Chandar, B., Muir, I., List, J. A., Woolridge, J. M., and Hortaçsu, A. (2019). Design and analysis of cluster-randomized field experiments in panel data settings. *Working paper*.

Charness, G. and Dufwenberg, M. (2006). Promises and Partnership. *Econometrica*, 74(6):1579–1601.

Charness, G. and Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics*, 117(3):817–869.

Chaudhry, S. and Loewenstein, G. (2017). Thanking, Apologizing, Bragging, and Blaming: The Currency of Communication.

Cohen, M. C., Fiszer, M. D., and Kim, B. J. (2018). Frustration-based Promotions: Field Experiments in Ride-Sharing. pages 1–42.

Cohen, P., Hahn, R., Hall, J., Levitt, S., and Metcalfe, R. (2016). Using Big Data to Estimate Consumer Surplus: The Case of Uber.

Cook, C., Diamond, R., Hall, J., List, J. A., and Oyer, P. (2018). The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers.

de Cremer, D., Pillutla, M. M., and Folmer, C. R. (2011). How important is an apology to you? Forecasting errors in evaluating the value of apologies. *Psychological Science*, 22(1):45–48.

Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47(2):268–298.

Fehr, E. and List, J. A. (2004). The Hidden Costs and Returns of IncentivesTrust and Trustworthiness among Ceos. *Journal of the European Economic Association*, 2(5):743–771.

Fischbacher, U. and Utikal, V. (2010). Learning and Peer Effects On the Acceptance of Apologies. (53).

Gilbert, B., James, A., and Shogren, J. (2017). Corporate Apology for Environmental Damage.

Hall, J. V., Horton, J. J., and Knoepfle, D. T. (2017). Labor Market Equilibration: Evidence from Uber. *Working Paper*, pages 1–42.

Harrison, G. W. and List, J. A. (2004). Field Experiments. *Journal of Economic Literature*.

Ho, B. (2012). Apologies as Signals: With Evidence from a Trust Game. *Management Science*, 58(1):141–158.

Ho, B. and Huffman, D. (2006). Trust and the Law. 12604(1977):1–29.

Ho, B. and Liu, E. (2011). Does sorry work? The impact of apology laws on medical malpractice. *Journal of Risk and Uncertainty*, 43(2):141–167.

Kim, P. H., Ferrin, D. L., Cooper, C. D., and Dirks, K. T. (2004). Removing the Shadow of Suspicion: The Effects of Apology Versus Denial for Repairing Competence- versus Integrity-Based Trust Violations. *Journal of Applied Psychology*, 89(1):104–118.

Levitt, S. D. and List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives*, 21(2):153–174.

Maddux, W. W., Kim, P. H., Okumura, T., and Brett, J. M. (2011). Cultural differences in the function and meaning of apologies. *International Negotiation*, 16(3):405–425.

Ohtsubo, Y., Watanabe, E., Kim, J., Kulas, J. T., Muluk, H., Nazar, G., Wang, F., and Zhang, J. (2012). Are costly apologies universally perceived as being sincere? *Journal of Evolutionary Psychology*, 10(4):187–204.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economic. *The American Economic Review*.

Schweitzer, M. E., Hershey, J. C., and Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes*, 101(1):1–19.

Sen, A. (1977). Rational Fools : A Critique of the Behavioral Foundations of Economic Theory. *Philosophy & Public Affairs*, 6(4):317–344.

Skarlicki, D. P., Folger, R., and Gee, J. (2004). When Social Accounts Backfire: The Exacerbating Effects of a Polite Message or an Apology on Reactions to an Unfair Outcome. *Journal of Applied Social Psychology*, 34(2):322–341.

# A Appendix A: Different Kinds of Apologies

A key part of the original design of the experiment was to test different kinds of apologies by modifying the text of the apology message. The intent was to identify evidence for the different mechanisms identified by Ho (2012), which classified apologies into one of five categories:

1. Costly apology: "I'm sorry, here's $5." An apology that involves a tangible cost.
2. Commitment apology: "I'm sorry, I won't do it again." An apology that promises to do better in the future. Based on a screening contract.
3. Status apology: "I'm sorry, I'm an idiot." An apology that admits incompetence. Based on two-dimensional type.
4. Empathy apology: "I'm sorry, I see that you are hurt." An apology that recognizes the other's pain. Based on information partitions.
5. Excuses: "I'm sorry, it wasn't my fault." An apology that blames external factors. Based on verifiable cheap talk.

Our study was designed to focus on the first three. Empathy was thought to be too difficult for a corporation to communicate over an email while excuses would have been technically more difficult and potentially had greater negative consequences.

The idea of the three types of apologies were conveyed to Uber's marketing department who designed messages consistent with the intent of the theory but also consistent with Uber's marketing practices.

## A.1 Commitment Apologies

The theoretical basis of the commitment apology is a screening contract. As noted in the main text, the principal (consumer) offers the agent (firm) a menu that says the following: If the firm apologizes for the breach of trust, then the relationship will be continued; however, if the firm is late again, the relationship will be immediately terminated in favor of the outside option.

In each round the principal has the option of staying with the current agent or choosing an outside option. In a commitment apology, the principal commits to a menu of rewarding the agent using its future decisions to stay with the current firm based on whether they apologized or not and their trustworthiness in future periods.

Table 5: Continuation values for commitment apologies.

| Agent Behavior | Cont. Value |
|---|---|
| Apologize, then good ride | $v_1^g$ |
| Apologize, then bad ride | $v_1^b$ |
| No apology, then good ride | $v_0^g$ |
| No apology, then bad ride | $v_0^b$ |

If good-intention firms are more likely to have good rides in the future, a separating equilibrium exists where good-intention firms apologize and accept the threat of immediate termination while bad-intention firms do not apologize and are judged in the future based on their performance (See Ho (2012) Online Appendix for details). The principal must commit to future behavior such that

$$v_1^g > v_0^g > v_0^b > v_0^b$$

. Note this is not renegotiation-proof since once a firm apologizes it reveals itself to be of good intentions. This suggests a role for emotional motivations that maintain the equilibrium behavior.

## A.2 Status Apologies

An alternate contract theory-inspired model for how apologies restore trust in a relationship is based on the idea that intrinsic type, $\theta$, is two-dimensional. A firm can have good intentions but they may be unreliable for some types of tasks (Chaudhry and Loewenstein (2017) provide recent evidence on how apologies rely on the trade-off in the the agent's perception of the principal's competence versus the principal's warmth). Suppose the distribution of external shocks, $\omega$ is correlated with the agent's type. The principal can choose which tasks to assign to the agent depending on her beliefs about the agent's type.

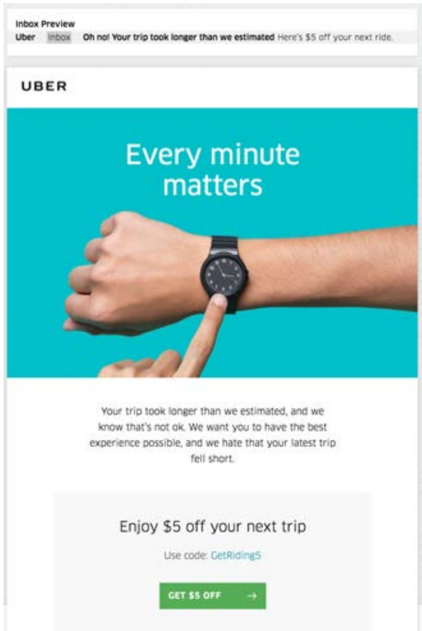
For example, suppose some firms are better suited for rides to the airport, while other firms are better suited for rides downtown. Here, the principal can offer a screening contract after a bad airport ride, where if the agent apologizes, they implicitly acknowledge their own inadequacy at airport rides, and if the agent doesn't apologize, they implicitly admit to having poor intentions. A separating equilibrium can be enforced if the agent can assign similar tasks in the future to agents who do not apologize but different tasks to agents who do apologize.

An agent with good intentions but who admits to being bad at providing airport rides will get more city rides in the future. An agent with bad intentions will not apologize because they find the airport rides to be more lucrative. The agent with good intentions would not choose to not apologize because they know they are bad at them. As in Battaglini (2002), the presence of multiple dimensions of type allows the principal to get fully revealing information from a cheap signal.

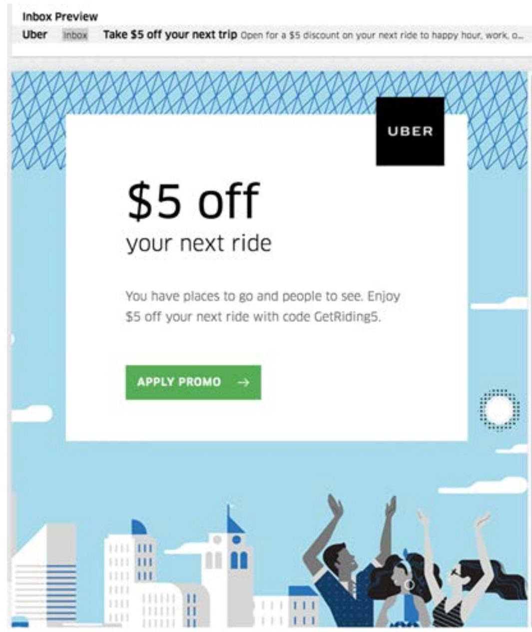
In our experiment we expected to find evidence for status-based apologies if a rider responded to an apology by decreasing future spending for similar rides but increasing it for dissimilar rides. However, testing for such effects for airport versus non-airport rides, weekday versus weekend rides, and rush hour versus non rush-hour rides, returned statistically indistinguishable treatment effects.

# B Appendix B: Experiment Details

As discussed in Appendix A, riders who were sent an apology email received one of three types – a "basic apology", a "status apology", or a "commitment apology" – that either included a \$5 promotion or did not. A screenshot of the basic apology email, with a promotion, is shown in Figure 10a. Additionally,




(a) Example screenshot of one of the apology emails sent to riders: the "basic" apology with a $5 promotion.




(b) Screenshot of the "just promo" email sent to riders, i.e. not including an explicit apology.

Figure 10

one treatment group received an email with the $5 promotion, but no explicit statement of apology, shown in Figure 10b. The apology emails differed in their subject lines and body paragraphs in the following way: Basic apology:

- Subject line: "Oh no! Your trip took longer than we estimated"
- Body: "Your trip took longer than we estimated, and we know that's not ok. We want you to have the best experience possible, and we hate that your latest trip fell short."

Commitment apology:

- Subject line: "We can do better."
- Body: "Your trip took longer than expected, and you deserve better.

This time we missed the mark, but we're working hard to give you arrival times that you can count on."

Status apology:

- Subject line: "We know our estimate was off."
- Body: "We underestimated how long your trip would take – and that's our fault. Every trip should be the best experience possible, and we recognize that your latest trip fell short."

As reported in the body of the paper, technological limitations meant that stratification could only be done for subjects who had signed up for the Uber platform before the start of the experiment. Subjects who joined after the start date were randomly assigned to one of the treatment groups. The same technological constraints meant that each treatment arm could only be allocated an integer percentage of the population. Since we have eight treatment arms, this meant that some treatment groups received 13% of newly registered users, while others received 12%. Because the newly registered users systematically differ from previously registered users, this resulted in the nonzero differences between our raw-means estimates and our regression-adjusted estimates. However, the differences do not impact our analysis and moreover, the results are similar if we drop all users who were not pre-randomized.